TIKHONOV REGULARIZATION

Stéphane Crépey
Département de Mathématiques
Université d'Évry Val d'Essonne
91025 Évry Cedex, France

September 19, 2008

Abstract

An important issue in quantitative finance is *model calibration*. The calibration problem is the *inverse* of the pricing problem. Instead of computing prices in a model with given values for its parameters, one wishes to compute the values of the model parameters that are consistent with observed prices. Now, it is well-known by physicists that such inverse problems are typically *ill-posed*. So, if one perturbs the data (e.g., if the observed prices move from some small amount between today and tomorrow), it is quite typical that a numerically determined best fit solution of the calibration problem switches from one 'basin of attraction' to the other, thus the numerically determined solution is*unstable*. To achieve robustness of model (re)calibration, we need to introduce some *regularization*. The most widely known and applicable regularization method is *Tikhonov(–Phillips)* regularization method. In this paper we provide a survey about Tikhonov regularization and we illustrate it by application to the problem of calibrating a local volatility model.

1 Financial Motivation

An important issue in quantitative finance is *model calibration*. The calibration problem is the *inverse* of the pricing problem. Instead of computing prices in a model with given values for its parameters, one wishes to compute the values of the model parameters that are consistent with observed prices (up to the bid–ask spread).

Now, it is well-known by physicists that such inverse problems are typically *ill-posed*. Recall that a problem is *well-posed* (as defined by Hadamard) if its solution exists, is unique, and depends continuously on its input data. Thus there are three reasons for which a problem might be ill-posed:
• it admits no solution, or/and
• it admits more than one solution, or/and
• the solution(s) of the inverse problem do(es) not depend on the input data in a continuous way.
In the case of calibration problems in finance, except for trivial situations, there exists typically *no instance* of a given class of models which is *exactly* consistent with a full calibration data set, including a number of option prices, a zero-coupons curve, an expected dividend yield curve on the underlying, etc. But there are often *various* instances of a given class of models that fit the data *within the bid–ask spread*. In this case, if one perturbs the data (e.g., if the observed prices move from some small amount between today and tomorrow), it is quite typical that a numerically determined best fit solution of the calibration problem switches from one 'basin of attraction' to the other, thus the numerically determined solution is *not stable* either.

In order to get a well-posed problem, we need to introduce some *regularization*. The most widely known and applicable regularization method is *Tikhonov(–Phillips)* regularization method [17, 15, 10].

2 Tikhonov regularization of non-linear inverse problems

We consider a Hilbert space $\mathcal{H}$, a closed convex non-void subset $\mathcal{A}$ of $\mathcal{H}$, a direct operator ('pricing functional')

$$\mathcal{H} \supseteq \mathcal{A} \ni a \xrightarrow{\Pi} \Pi(a) \in \mathbb{R}^d ,$$

(so a corresponds to the set of model parameters), noisy data ('observed prices') π^δ, and a *prior* $a_0 \in \mathcal{H}$ (a priori guess for a). The Tikhonov regularization method for *inverting* Π *at* π^δ, or estimating the model parameter a given the observation π^δ, consists in:
• reformulating the inverse problem as the following *nonlinear least squares problem*:

$$\min_{a \in \mathcal{A}} \left\| \Pi(a) - \pi^\delta \right\|^2 \tag{1}$$

to ensure *existence* of a solution,
• selecting the solutions of the previous nonlinear least squares problem that minimize $\|a - a_0\|^2$ over the set of all solutions, and
• introducing a trade-off between accuracy and regularity, parameterized by a level of regularization $\alpha > 0$, to ensure *stability.*
More precisely, we introduce the following *cost criterion*:

$$J_\alpha^\delta(a) \equiv \left\| \Pi(a) - \pi^\delta \right\|^2 + \alpha \|a - a_0\|^2 \ . \tag{2}$$

Given α, δ and a further parameter η, where η represents an error tolerance on the minimization, we define a *regularized solution to the inverse problem for* Π *at* π^δ, as any model parameter $a_\alpha^{\delta,\eta} \in \mathcal{A}$ such that

$$J_\alpha^\delta\left(a_\alpha^{\delta,\eta}\right) \leq J_\alpha^\delta(a) + \eta \quad , \quad a \in \mathcal{A} \ .$$

Under suitable assumptions, one can show that the regularized inverse problem is well-posed, as follows. We first postulate that the direct operator Π satisfies the following regularity assumption.

Assumption 2.1 (Compactness) $\Pi(a_n)$ converges to $\Pi(a)$ in $\mathbb{R}^d$ if a_n weakly-converges to a in $\mathcal{H}$.

We then have the following *stability* result.

Theorem 2.1 (Stability) *Let* $\pi^{\delta_n} \to \pi^\delta$, $\eta_n \to 0$ *when* $n \to \infty$. *Then any sequence of regularized solutions* $a_\alpha^{\delta_n,\eta_n}$ *admits a subsequence which converges towards a regularized solution* $a_\alpha^{\delta,\eta=0}$.

Assuming further that the data lie in the range of the model leads to *convergence* properties of regularized solutions to (unregularized) solutions of the inverse problem as $\alpha \to 0$. Let us then make the following additional assumption on Π.

Assumption 2.2 (Range property) $\pi \in \Pi(\mathcal{A})$.

By an a_0 – *solution* to the inverse problem for Π at π, we mean any $a \in \operatorname{Argmin}_{\{\Pi(a)=\pi\}} \|a - a_0\|$. Note that the set of a_0-solutions is non-empty, by Assumption 2.2.

Theorem 2.2 (Convergence; see, for instance, Theorem 2.3 of Engl *et al* [11]) *Let the* perturbed parameters $\alpha_n, \delta_n, \eta_n$ *and the* perturbed data $\pi_n \in \mathbb{R}^d$ *satisfy*

$$(n \in \mathbb{N}) \quad \|\pi - \pi_n\| \leq \delta_n,$$

$$(n \to \infty) \quad \alpha_n \quad , \quad \delta_n^2/\alpha_n \quad , \quad \eta_n/\alpha_n \quad \longrightarrow \quad 0.$$

Then any sequence of regularized solutions $a_{\alpha_n}^{\delta_n,\eta_n}$ admits a subsequence which converges towards an a_0-solution a of the inverse problem for Π at π. In particular, in case when this problem admits a unique *a_0-solution a, then $a_{\alpha_n}^{\delta_n,\eta_n}$ converges to a.*

Remark 2.3 In the special case where the direct operator Π is linear, Tikhonov regularization thus appears as an approximating scheme for the pseudo-inverse of Π.

Finally, assuming further regularity of Π, one can get *convergence rates* estimates, uniform over all data $\pi \in \Pi(\mathcal{A})$ sufficiently close and smooth with respect to the prior a_0 (so that the additional *source condition* (3) is satisfied). Let us thus make the following additional assumption on Π.

Assumption 2.4 (Twice Gateaux differentiability) There exists linear and bilinear forms $d\Pi(a)$ on $\mathcal{H}$ and $d^2\Pi(a)$ on $\mathcal{H}^2$ such that

$$\Pi(a+\varepsilon h) = \Pi(a) + \varepsilon d\Pi(a)\cdot h + \tfrac{\varepsilon^2}{2} d^2\Pi(a)\cdot(h,h) + \mathrm{o}\left(\varepsilon^2\right) \quad ; \quad a, a+h \in \mathcal{A}$$
$$\|d\Pi(a)\cdot h\| \le C\|h\| \quad , \quad \left\|d^2\Pi(a)\cdot(h,h')\right\| \le C\|h\|\|h'\| \quad ; \quad a \in \mathcal{A} \ , \quad h,h' \in \mathcal{H}$$

where C is a constant independant of $a \in \mathcal{A}$.

In the following theorem the operator

$$d\Pi(a)^* : \mathbb{R}^d \ni \lambda \mapsto d\Pi(a)^*\lambda \in \mathcal{H}^1$$

denotes the *adjoint* of

$$d\Pi(a) : \mathcal{H}^1 \ni h \mapsto d\Pi(a)h \in \mathbb{R}^d \, ,$$

in the sense that (see [10]):

$$\langle h, d\Pi(a)^\star\lambda\rangle_{\mathcal{H}^1} \;=\; \lambda' d\Pi(a).h \;; \quad (h,\lambda) \in \mathcal{H}^1 \times \mathbb{R}^d \, .$$

Theorem 2.3 (Convergence Rates; see, for instance, Theorem 10.4 of Engl *et al* [10]) *Assume*

$$(n \in \mathbb{N}) \quad \|\pi - \pi_n\| \le \delta_n,$$

$$(n \to \infty) \quad \alpha_n \longrightarrow 0 \quad , \quad \alpha_n \sim \delta_n \quad , \quad \eta_n = \mathrm{O}\left(\delta_n^2\right) \, .$$

Then $\|a_{\alpha_n}^{\delta_n,\eta_n} - a\| = \mathrm{O}(\sqrt{\delta_n})$, for any a_0-solution a of the inverse problem for Π at π such that

$$a - a_0 = d\Pi(a)^*\lambda \tag{3}$$

for some λ sufficiently small in $\mathbb{R}^d$ (in particular, there exists at most one such a_0-solution a).

Remark 2.5 An interesting feature of Tikhonov regularization is that the data set π does not need to belong to the range of the direct operator for applicability of the method — even if Assumption 2.2 is the simplest assumption for the previous results regarding convergence and convergence rates (in fact a minimal assumption for such results is the existence of a least squares solution to the inverse problem, see Proposition 3.2 of Binder *et al* [2]).

An important issue in practice is the choice of the *regularization parameter* α, that determines the trade-off between accuracy and regularity in the method. To set α, the two main approaches are:
- *a priori* methods, in which the choice of α only depends on δ, the level of noise on the data (such as the size of the bid–ask spread, in the case of market prices data in finance);
- more general *a posteriori* methods, in which α may depend on the data in a less specific way.

In applications to calibration problems in finance, the most commonly used method for choosing α is the a posteriori method based on the so-called *discrepancy principle*, which consists in choosing the greatest level of α for which the 'distance' $\left\|\Pi\left(a_\alpha^{\delta,\eta}\right) - \pi^\delta\right\|$ (for given δ,η) does not exceed the level of noise δ on the observations (as measured by the bid-ask spread).

2.1 Implementation

For implementation purposes, the minimization problem (2) is discretized, thus becoming effectively a *nonlinear minimization problem* on (some subset of) $\mathbb{R}^k$ (see, e.g., [14]), where k is the number of model parameters to be estimated.

In the case of a *strictly convex* cost criterion $J = J_\alpha^\delta$ in (2), and if, additionally, J is differentiable, one can prove the convergence to the (unique) minimum of various *gradient descent algorithms*. These consist in moving at each step from some amount (fixed step descent *vs* optimal step descent) in a direction defined by the gradient ∇J at the current step of the algorithm, in combination with, in some variants of the method *(conjugate gradient method, quasi-Newton algorithms, etc)*, the gradient(s) ∇J at the previous step(s).

In the *non strictly convex* case, (actually, in the context of calibration problems in finance, J is typically not even convex w.r.t. a), or if the cost criterion is only almost everywhere differentiable (as in the *American calibration problem*, see Remark 3.1(i)), such algorithms can still be used, in which case they typically converge to one among many *local minima* of J.
When there are no constraints (case $\mathcal{A} = \mathcal{H}$), the minimization problem is, in practice, much easier, and many implementations of the related gradient descent algorithms are available (see for instance [16]). As for constrained problems, a state-of-the-art open-source implementation of the quasi-Newton method for minimizing a function on a box, the lbfgs algorithm, is available on `www.ece.northwestern.edu/~nocedal/lbfgsb.html`.

When the gradient ∇J is not computable in closed form, and not computable numerically with the required accuracy either, an alternative to gradient descent methods is to use the *nonlinear simplex method* (not to be confused with the simplex algorithm for solving linear programming problems, see [16]). As opposed to gradient descent methods, the nonlinear simplex algorithm only uses the *values* (and not the *gradient*) of J, but the convergence of the algorithm is not proved in general, and there are known counter-examples in which it does not converge.

3 Application: Extracting Local Volatility

In the case of *parametric* models in finance, namely models with a small number of *scalar* parameters, such as Heston's stochastic volatility model or Merton's jump-diffusion model (as opposed to models with *functional*, e.g., time-dependent, parameters), the choice of a suitable regularization term is generally not obvious. In this case, the calibration industry standard rather consists in solving the unregularized non linear least squares problem (1). So Tikhonov regularization is rather used for calibrating *non parametric* financial models.

In this Section we consider the problem of inferring a *local volatility function* $\sigma(t, S)$ (see Dupire [8]) from observed option prices, namely European vanilla calls and/or puts with various strikes and maturities on the underlying S. The local volatility function thus inferred may then be used to price exotic options and/or Greeking, consistently with the market (see, for instance, Crépey [6]).

3.1 The ill-posed Local Volatility Calibration problem

But the local volatility calibration problem is under-determined (since the set of observed prices is finite whereas the nonparametric function σ has an infinity of degrees of freedom) and ill-posed. So a naïve approach based on numerical differentiation using the so-called *Dupire's formula* [8] gives a local volatility which is highly oscillatory (see Figure 1), and thus unstable, for instance when performing a day-to-day calibration.

To meet this issue, the first idea that comes to mind is to seek for σ within a parameterized family of functions. However finding classes of functions with all the flexibility required for fitting implied volatility surfaces with several hundred of implied volatility points and a variety of shapes, turns out to be a very challenging task (unless a large family of splines is considered, see Coleman et al. [3], in which case the ill-posedness of the problem shows up again).

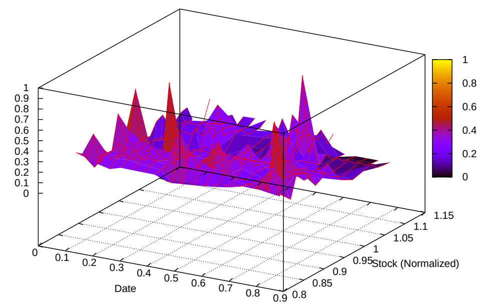

Figure 1: *Local Variance $\sigma(t,S)^2$ obtained by application of Dupire's formula on the DAX index, May 2 2001.*

The best way to proceed is to stay non-parametric, and to use regularization methods to stabilize the calibration procedure. Since we use a non-parametric local volatility, the model contains a sufficient number of degrees of freedom to provide a perfect fit to virtually any market smile. And the regularization method guarantees that the local volatility thus calibrated is nice and smooth.

3.2 Approach by Tikhonov regularization

Among the various regularization methods at hand, the most popular one is the Tikhonov regularization method of Section 2. One thus rewrites the local volatility calibration problem as the following nonlinear minimization problem:

$$\min_{\{\sigma\equiv\sigma(t,S);\underline{\sigma}\le\sigma\le\overline{\sigma}\}} J(\sigma) = \|\Pi(\sigma)-\pi\|^2 + \alpha\|\sigma-\sigma_0\|^2_{\mathcal{H}^1} \tag{4}$$

where:
• the bounds $\underline{\sigma}$ and $\overline{\sigma}$ are given positive constants specifying the abstract set $\mathcal{A}$ of Section 2,
• π is the vector of market prices observed at the calibration time,
• $\Pi(\sigma)$ is the related vector of prices in the Dupire model with volatility function σ,
• σ_0 is a suitable prior (a priori guess on σ), and for $u \equiv u(t,S)$:

$$\|u\|^2_{\mathcal{H}^1} := \int_{t_0}^{\infty}\int_0^{\infty}\left[u(t,S)^2 + (\partial_t u(t,S))^2 + (\partial_S u(t,S))^2\right] dtdS\ .$$

Problem (4) and a related gradient descent approach to solve it numerically (cf. Subsection 2.1) were introduced in Lagnado and Osher [13]. Crépey [7] (see also Egger and Engl [9]) further showed that the general conditions of Section 2 are satisfied in this case. Stability and convergence of the method follow.

In Crépey [6] an efficient trinomial tree implementation of this approach was presented, based on an exact computation of the gradient of the (discretized) cost criterion J in (4). Figure 2 displays the local variance surface $\sigma(t,S)^2$ (to be compared with that of Figure 1), the corresponding implied volatility surface and the accuracy of the calibration, obtained by running this algorithm on the DAX index European options data set of May 2, 2001 (consisting of about 300 European vanilla option prices distributed throughout 6 maturities with moneyness $K/S_0 \in [0.8, 1.2]$). At the initiation of the algorithm, the norm of the gradient of the cost criterion J in (4) was equal to `5.73E-02`, and upon convergence after 65 iterations of the gradient descent algorithm, a local minimum of the cost criterion was found, with related value of the norm of the gradient of the cost criterion equal to `6.83E-07`. In the accuracy graph, `implied volatility mismatch` refers to the difference between the Black–Scholes implied volatility corresponding to the market price of an option and its price in

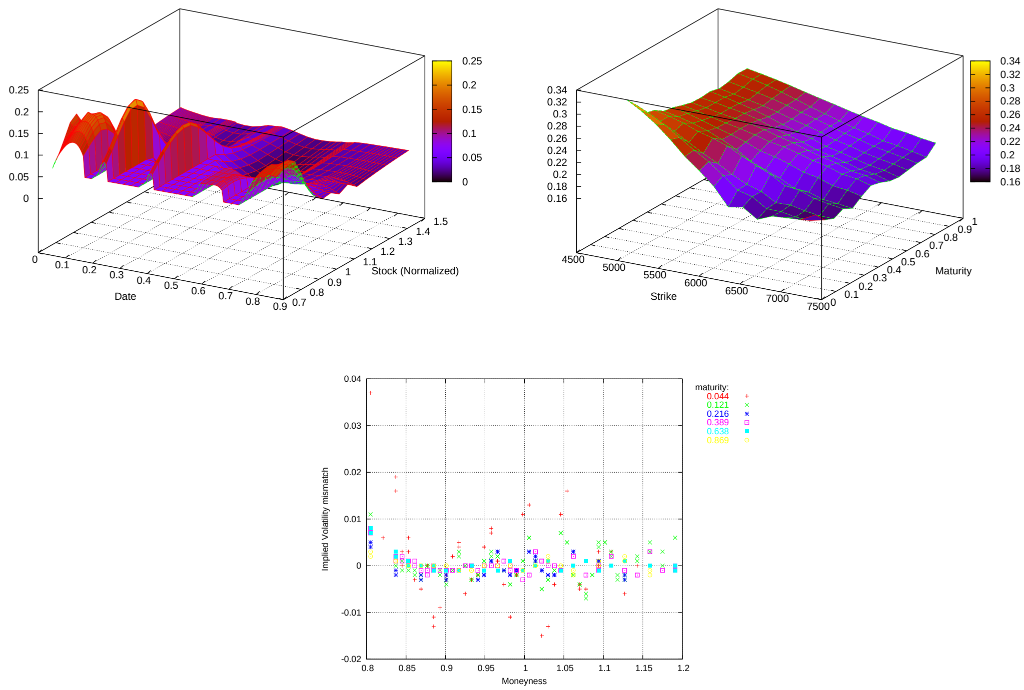

Figure 2: *Local variance, implied volatility and calibration accuracy obtained by application of the Tikhonov regularization method on the DAX index, May 2 2001.*

the calibrated local volatility model, for each option in the calibration data set.

Such calibration procedures are typically computationally intensive, however it is possible to make them faster by resorting to *parallel computing* (see Table 1 and Crépey [6]).

n × nproc	**1**	**3**	**6**
54	25s	9s	10s
101	4m30s	1m57s	1m36s

Table 1: *Calibration CPU times on a cluster of* nproc *1.3 GHz processors connected on a fast Myrinet network, using a calibration tree with n time steps (thus $n^2/2$ nodes in the tree).*

Remark 3.1 (i) This approach by Tikhonov regularization can be extended to the problem of calibrating a local volatility function using *American* observed option prices as input data (see Crépey [6]), or to the problem of calibrating a *Lévy model with local jump measure* (see Cont and Rouis [4], Kindermann et al. [12]).
(ii) An alternative approach for this problem is to use *entropic regularization*, rewriting the local volatility calibration problem as a related *stochastic control problem* (see Avellaneda et al. [1]).

References

[1] M. AVELLANEDA, C. FRIEDMAN, R. HOLMES AND D. SAMPERI. Calibrating volatility surfaces via relative-entropy minimization, *Applied Math. Finance*, 41 (1997), pp. 37–64.

[2] A. BINDER, H. W. ENGL, C. W. GROETSCH, A. NEUBAUER, AND O. SCHERZER. Weakly closed nonlinear operators and parameter identification in parabolic equations by Tikhonov regularization, *Appl. Anal.* 55 (1994), pp. 13–25.

[3] T. COLEMAN, Y. LI AND A. VERMA. Reconstructing the unknown volatility function. *Journal of Computational Finance*, 2 (1999), 3, pp. 77–102.

[4] R. CONT AND M. ROUIS. Estimating exponential Lévy models from option prices via Tikhonov regularization, *Working Paper.*

[5] S. CRÉPEY. Delta-hedging Vega Risk? *Quantitative Finance*, 4 (October 2004), pp. 559–579.

[6] S. CRÉPEY. Calibration of the Local Volatility in a trinomial tree using Tikhonov regularization. *Inverse Problems*, 19 (2003), pp. 91–127.

[7] S. CRÉPEY. Calibration of the Local Volatility in a generalized Black–Scholes model using Tikhonov regularization. *SIAM Journal on Mathematical Analysis*, Vol. 34 No 5 (2003), pp. 1183–1206.

[8] B. DUPIRE. Pricing with a smile, *Risk*, 7 (1994), pp. 18–20.

[9] H. EGGER AND H. W. ENGL. Tikhonov Regularization Applied to the Inverse Problem of Option Pricing: Convergence Analysis and Rates, *Inverse Problems*, 21 (2005) 1027-1045

[10] H. W. ENGL, M. HANKE, AND A. NEUBAUER. *Regularization of Inverse Problems*. Kluwer, Dordrecht, 1996.

[11] H. W. ENGL, K. KUNISCH, AND A. NEUBAUER. Convergence rates for Tikhonov regularisation of nonlinear ill-posed problems. *Inverse Problems*, 5 (1989), 4, pp. 523–540.

[12] S. KINDERMANN, P. MAYER, H. ALBRECHER, H. W. ENGL. Identification of the local speed function in a Lévy model for option pricing, *Submitted*.

[13] R. LAGNADO AND S. OSHER. A Technique for Calibrating Derivative Security Pricing Models: Numerical Solution of an Inverse Problem, *Journal of Computational Finance*, 1 (1997), 1, pp. 13–25.

[14] J. NOCEDAL, S. J. WRIGHT. *Numerical Optimization, 2^{nd} edition*, *Springer*, 2006.

[15] D. PHILLIPS. A technique for the numerical solution of certain integral equations of the first kind, *J Assoc Comput Mach*, 9, 84-97 (1962).

[16] W.H. PRESS, B.P. FLANNERY, S.A. TEUKOLSKY AND W.T. VETTERLING, Numerical Recipes in C: The Art of Scientific Computing (Second Edition), *Cambridge University Press*, 1992.

[17] A. TIKHONOV. Solution of incorrectly formulated problems and the regularization method, *Soviet Math Dokl* 4, 1035-1038, English translation of *Dokl Akad Nauk SSSR* 151, 501-504, 1963.